

**REMARKS**

The fact that April 9, 2006, fell on a Sunday ensures that this paper is timely filed as of Monday, April 10, 2006, the next business day.

In the Office Action dated February 9, 2006, pending Claims 1-19 were rejected and the rejection made final. In response Applicants have filed herewith an Amendment After Final and respectfully request the Office to reconsider the rejections presented in the outstanding Office Action in light of the following remarks.

Before addressing the rejections on the merits, Applications would like to first address the objections to the specification in the outstanding Office Action. Reconsideration and withdrawal of these objections is respectfully requested in view of the following comments.

The first objection is addressed towards language appearing on Page 6, line 7. The language in contention is objected to because of a concern that the usage of the letter "N" is unclear in certain areas of the specification. In some appearances, the letter "N" is in italics; in some appearances, it is not formatted in any way. Regardless of format, the letter maintains the same definition throughout the specification. When used in formulas, the letter is italicized because of formatting issues. When creating a formula in a program such as Microsoft Word, the letters in the formula are automatically italicized. The use of such italics was not intended by Applicants to distinguish or differentiate the letter, but was rather a consequence of using the letter in a formula. Thus, the rejection is

Atty. Docket No. JP920000191US1  
(590.079)

respectfully traversed. It is respectfully requested that this objection be reconsidered and withdrawn.

Finally, the specification is objected to because the term 'ANWE' allegedly lacks antecedent definition or description. In accordance with the Examiner's suggestion, the specification has been amended to address this issue. Reconsideration and withdrawal of this objection is respectfully requested.

Claims 1-19 were pending in the instant application at the time of the outstanding Office Action. Of these claims, Claims 1, 10, and 19 are independent claims; the remaining claims are dependent claims.

Claims 1-3, 6-12, and 15-19 stand rejected under 35 USC § 103(a) as being unpatentable over Wang et al. (hereinafter "Wang") in view of Razin et al. (hereinafter "Razin"). Reconsideration and withdrawal of the present rejections are hereby respectfully requested.

The present invention is directed to a method and apparatus for automatically extracting new words from a cleaned corpus. The instant invention segments a cleaned corpus to form a segmented corpus, splits the segmented corpus to form sub strings, and counts the occurrences of each sub strings appearing in the given corpus. Finally, the present invention filters out false candidates to output new words.

As best understood, Wang appears to be directed to a method that optimizes language models in which an initial language model is developed from a lexicon and segmentation derived from a received corpus. The initial model is iteratively refined by

Atty. Docket No. JP920000191US1  
(590.079)

updating the lexicon and re-segmenting the corpus using both maximum match techniques and statistical principles. (Abstract) As asserted in the outstanding Office Action, Wang does not expressly disclose filtering out false candidates to output new words.

However, it is asserted that Razin discloses filtering out false candidates to output new words. As best understood, Razin appears to be directed to standardizing phrasing in a document. Razin identifies phrases in a document to create a preliminary list of phrases, then filters and refines those phrases to create a final list of standard phrases. Razin then identifies phrase of a document that are similar to standard phrases, decides if the candidate phrase is similar enough to the standard phrase and compute phrase substitutions to determine the approximate conformation of the standard phrase to the approximate phrase and vice versa. (Abstract)

The Office asserts that a phrase consists of words, and thus the ability of Razin to output new phrases in a document is equivalent to an ability to output new words. However, Applicant respectfully disagrees with this assertion. A new word consists of a unique and novel arrangement of character strings that did not previously exist in a language and which cannot be found in dictionaries. A new phrase is a new arrangement of words (that exist in dictionaries and are used in the language).

Razin explicitly asserts that his invention uses a tree based on stemmed words and known elements, not character strings, as in the instant invention. (Col. 1, line 54 to Col. 2, line 2) Razin stems words before creating a tree or any other structure in order to find

new phrases. Thus, a new word would not even be considered in its original form.

Rather, according to Razin, the word would be taken either as a misspelling or a conjugation, or some other form of a base word to which it can be “conformed”. (Col. 4, lines 18-26; Col. 6, lines 5-12) Thus, Razin actually teaches away from outputting or filtering new words in favor of finding new configurations of known words to identify as new phrases. Thus, it is respectfully submitted that there is no teaching or suggestion in Razin to filter out false candidates in order to output new words.

Claim 1 recites, *inter alia*, filtering out false candidates **to output new words**. (emphasis added) Similar language also appears in the other Independent Claims. Neither Wang nor Razin, nor the combination of the two, teach or suggest the limitations of the instant invention.

Further, a 35 USC 103(a) rejection requires that the combined cited references provide both the motivation to combine the references and an expectation of success. Not only is there no motivation to combine the references, no expectation of success, but actually combining the references would not produce the claimed invention. Thus, the claimed invention is patentable over the combined references and the state of the art.

There is an inherent tension in Wang and Razin given that, in Wang, the invention deals with word-level trees and does not relate at all to the process of standardized document phrasing, which is the crux of the invention of Razin. In fact, Razin specifically references this tension in a document similar to Wang and teaches away from combining the two references. (Col. 2, lines 22-40) Additionally, Razin asserts that his

Atty. Docket No. JP920000191US1  
(590.079)

invention uses a tree based on stemmed words and known elements, not character strings, as Wang. (Col. 1, line 54 to Col. 2, line 2) At stated before, however, combining Wang and Razin would result in producing a language model of phrases which includes a lexicon of standard phrases and known words. Even if there were a motivation for the combination, this combination does not teach or suggest the claimed invention.

Claims 4-5 and 13-14 stand rejected under 35 USC § 103(a) as being unpatentable over Wang et al. (hereinafter “Wang”) in view of Razin et al. (hereinafter “Razin”) and further in view of Hui. Specifically the Office asserted that “[i]t would have been obvious ... to modify Wang in view of Razin by specifically providing using extended suffix tree (GST or GAST), for the purpose of storing more than one input strings.” Reconsideration and withdrawal of this rejection is hereby respectfully requested.

Hui does not overcome the deficiencies of Wang or Razin. As best understood, Hui is directed towards an algorithm that provides an optimal sequential solution of the color set size problem which entails finding the number of different leaf colors in a subtree rooted at a vertex v in a rooted tree. Although Hui asserts that there is applicability in string matching heuristics, there is no teaching or suggestion in Hui to filter false word candidates to output new words.

Combining Wang, Razin, and Hui would result in producing a language model of phrases using an optimal sequential solution to find the phrases that constitute the lexicon of standard phrases. Even if there were a motivation for the combination, this combination does not teach or suggest the claimed invention.

04-10-'06 20:53 FROM-

412-741-9292

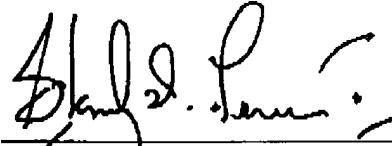
T-520 P016/016 F-689

Atty. Docket No. JP920000191US1  
(590.079)

In view of the foregoing, it is respectfully submitted that Independent Claims 1, 10 and 19 fully distinguish over the applied art and are thus allowable. By virtue of dependence from Claims 1 and 10, it is thus also submitted that Claims 2-9 and 11-18 are also allowable at this juncture.

In summary, it is respectfully submitted that the instant application, including Claims 1-19, is presently in condition for allowance. Notice to the effect is hereby earnestly solicited. If there are any further issues in this application, the Examiner is invited to contact the undersigned at the telephone number listed below.

Respectfully submitted,



Stanley D. Ference III  
Registration No. 33,879

**Customer No. 35195**  
FERENCE & ASSOCIATES  
409 Broad Street  
Pittsburgh, Pennsylvania 15143  
(412) 741-8400  
(412) 741-9292 - Facsimile

Attorneys for Applicants